SEARCHES FOR MASS-ASYMMETRIC COMPACT BINARY COALESCENCE EVENTS USING CONVOLUTIONAL NEURAL NETWORKS

Marc Andrés-Carcasona[1], A. Menéndez-Vázquez[1], M. Martínez[1,2], LL. M. Mir[1]

[1] Institut de Física d'Altes Energies (IFAE), Barcelona Institute of Science and Technology, E-08193 Barcelona, Spain

[2] Catalan Institution for Research and Advanced Studies (ICREA), E-08010 Barcelona, Spain

Mail to mandres@ifae.es

G2NET-WG1 meeting – 08/04/2022

Outline

1 Introduction

- Current methods
- Potential improvement of machine learning

2 Data Preparation

- Parameter space covered
- Generation of training sets

3 CNN Architecture

- Description of the layers that form the CNN
- Training hyper-parameters

4 Training Results

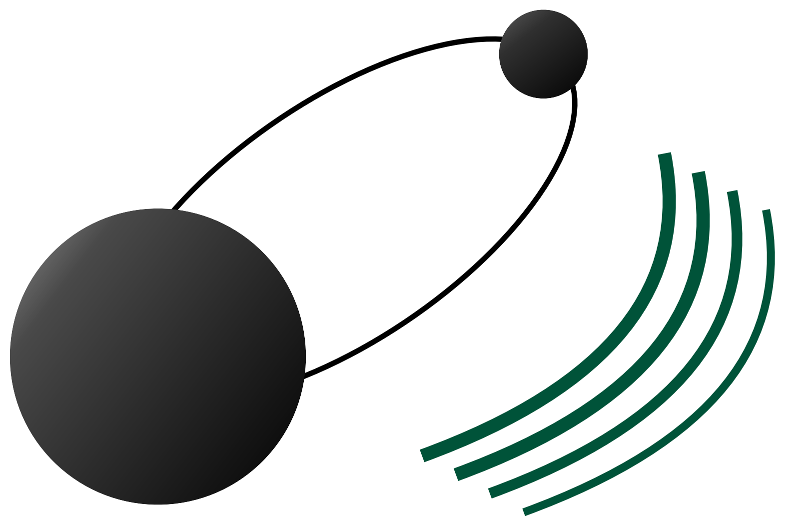
- Metrics of the training procedure
- Determination of the threshold
- Combination of outputs

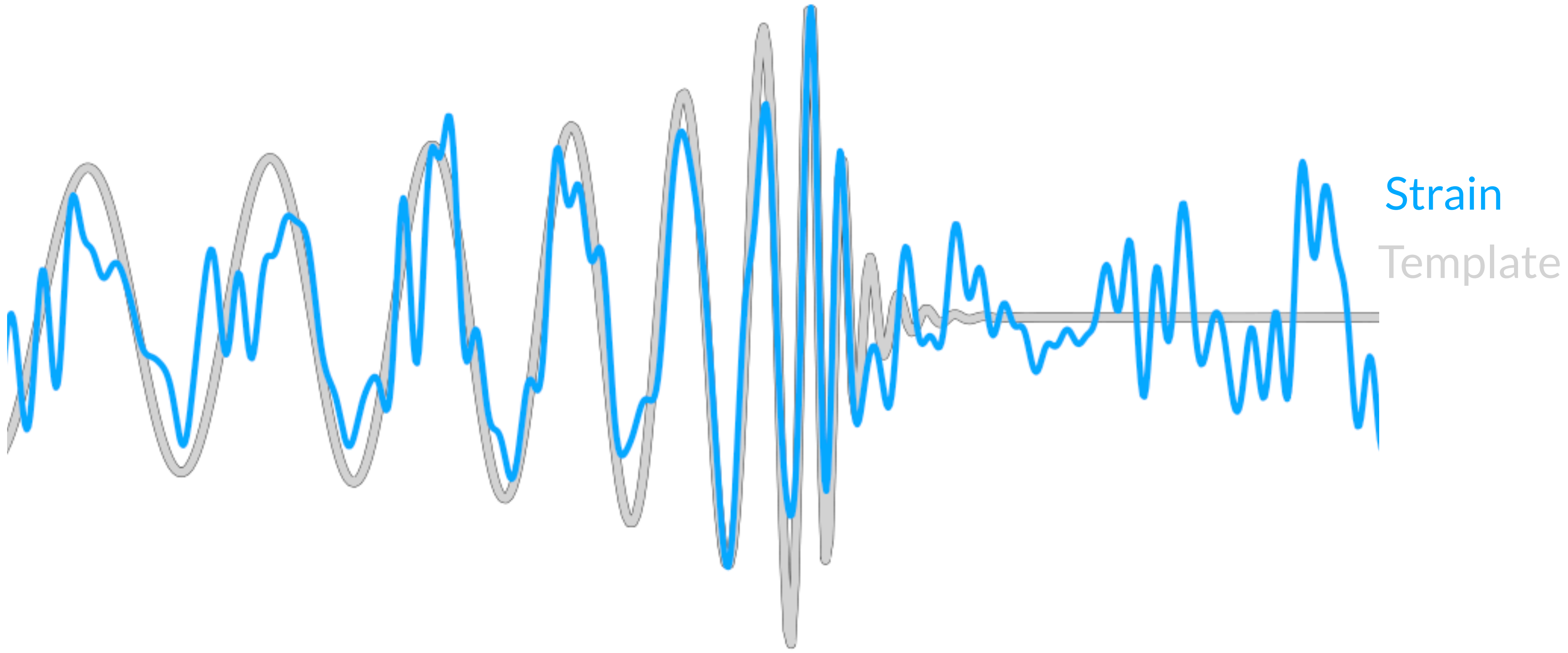
5 O3 Scan Results

- Events detected by the CNNs
- Comparison with GWTC-3 catalog

INTRODUCTION

We are interested in Compact Binary Coalescences with high mass ratio.



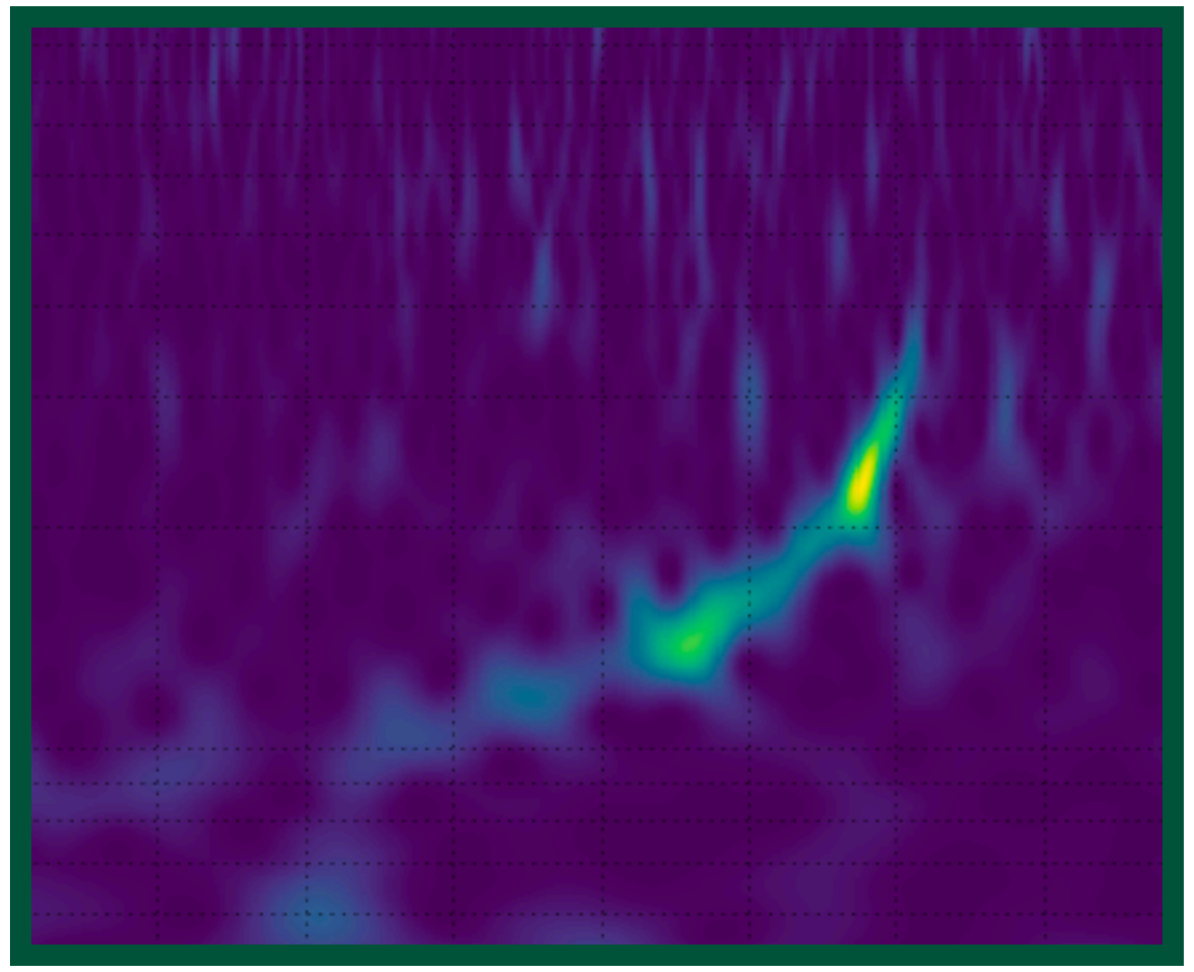


The traditional searching method is based on a matched filtering technique applied over a complete template bank. This method is robust but slow.

Alternatively, we can use a spectrogram to visualize the data and this two-dimensional image is perfect to be used alongside image processing Machine Learning (ML) techniques.

Some benefits of using ML are:

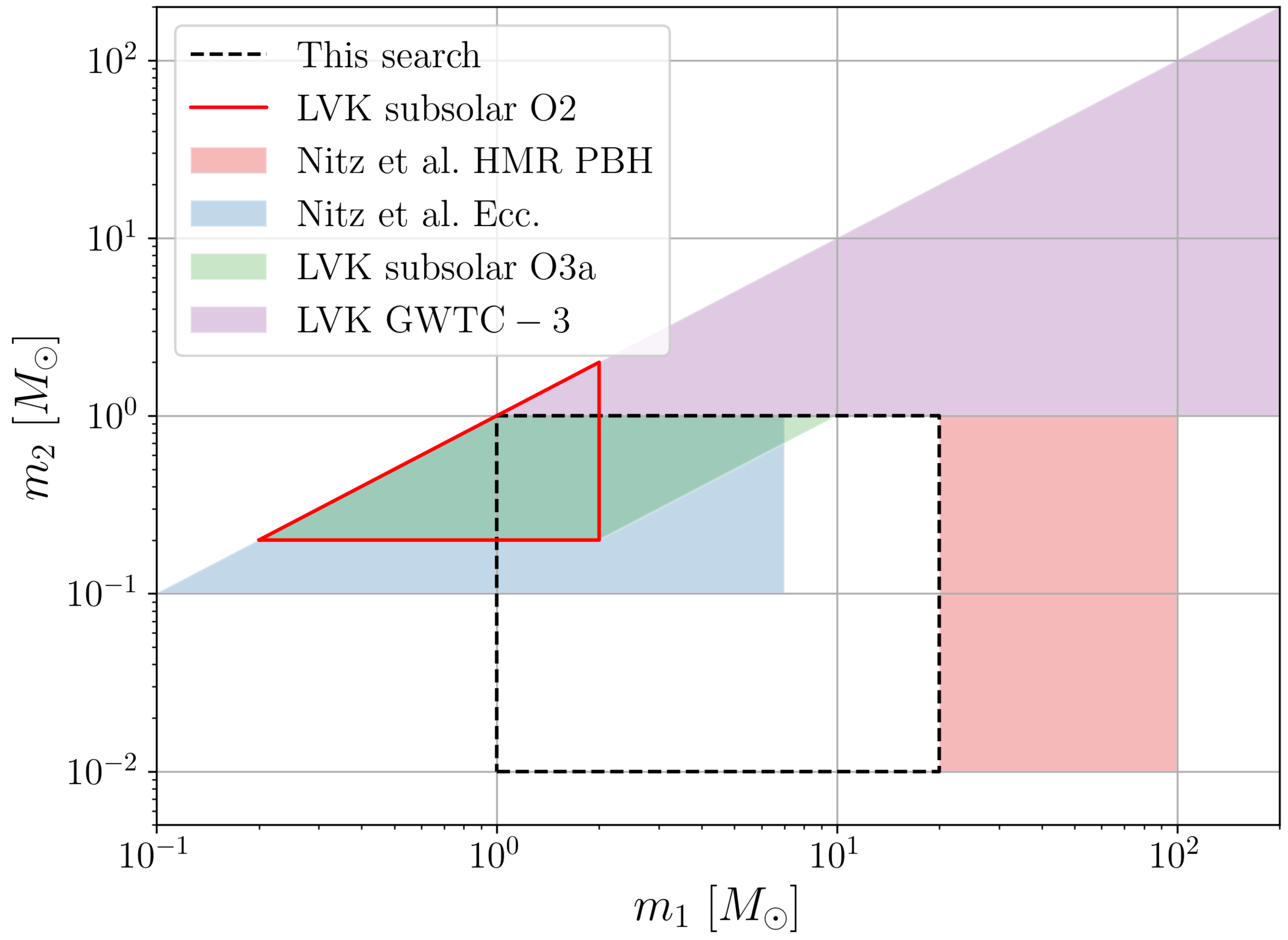
- Extensive literature from other fields
- Fast inference speeds
- Easy implementation due to preexisting libraries



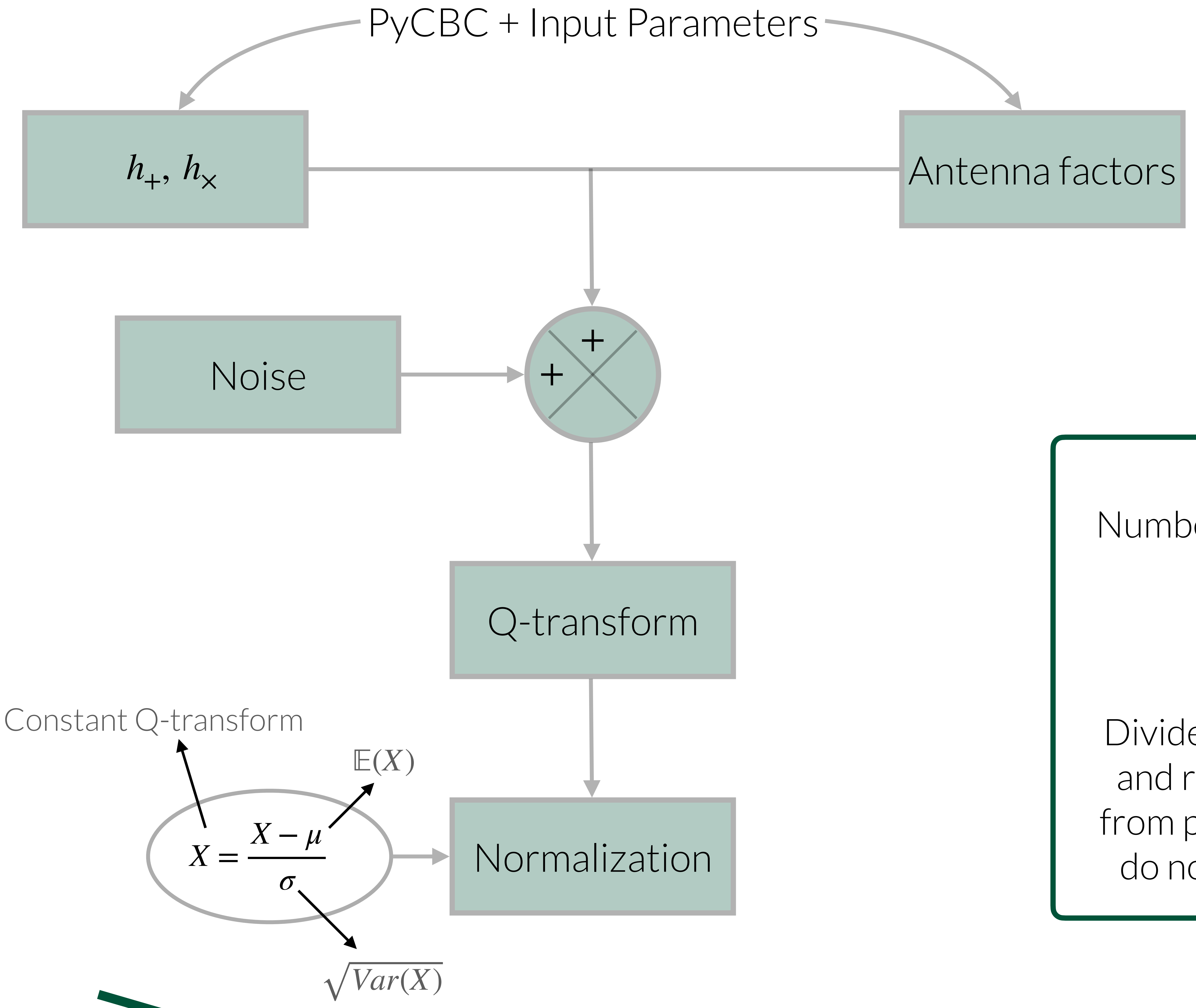
Inclination	Polarization
$\psi \in [0,\ \pi/2]$	$\theta_{JN} \in [0,\ \pi]$
Right Ascension	Declination
$\alpha \in [0,\ 2\pi]$	$\delta \in [-\pi/2,\ \pi/2]$

Masses $m_1 \in [1,\ 20]\ M_{\odot}$, $m_2 \in [0.01,\ 1]\ M_{\odot}$ $\longrightarrow$ $q \in [1,\ 2000]$

$$D_L \in [1,\ 100]\ Mpc$$



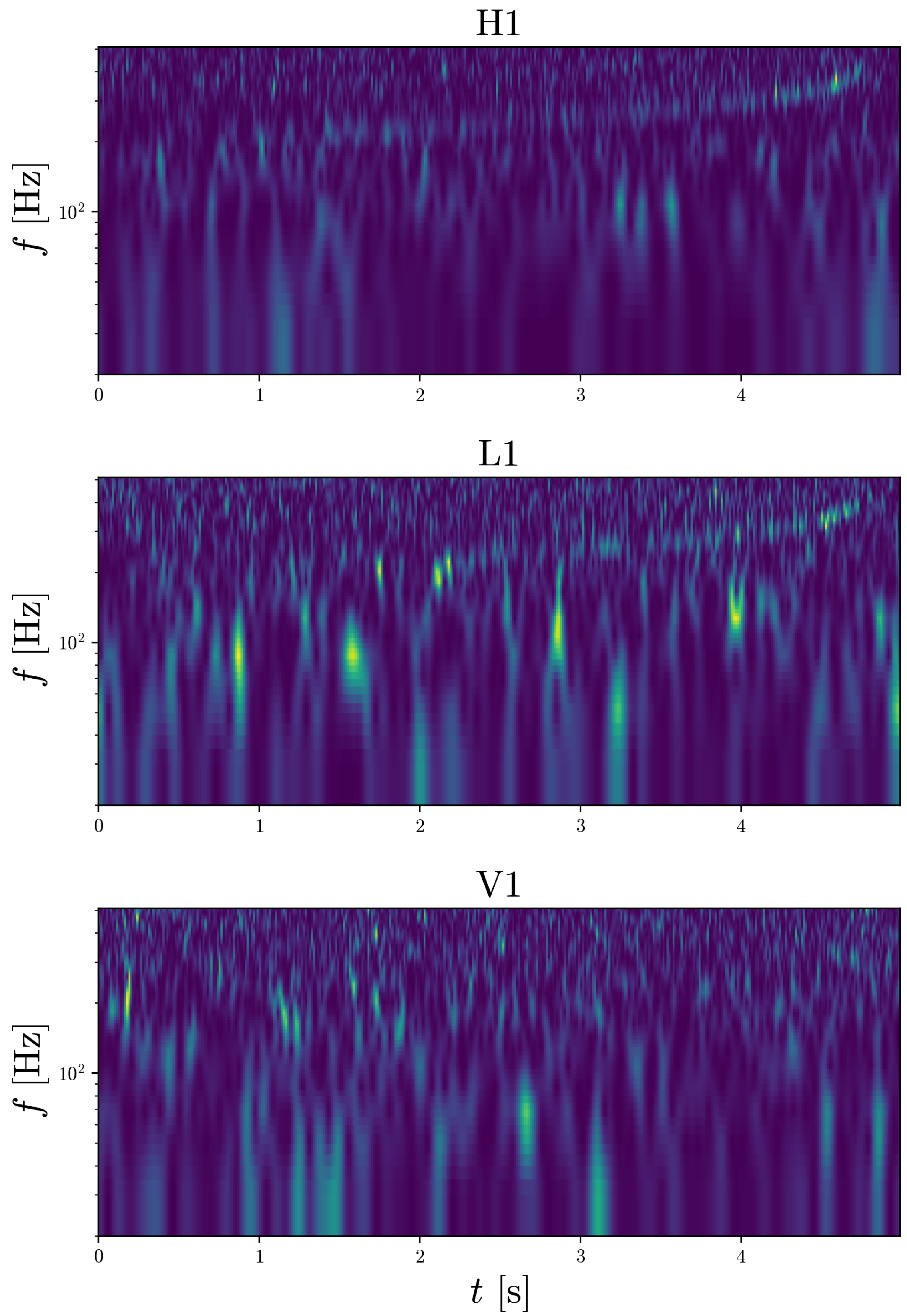
DATA PREPARATION - Images generation



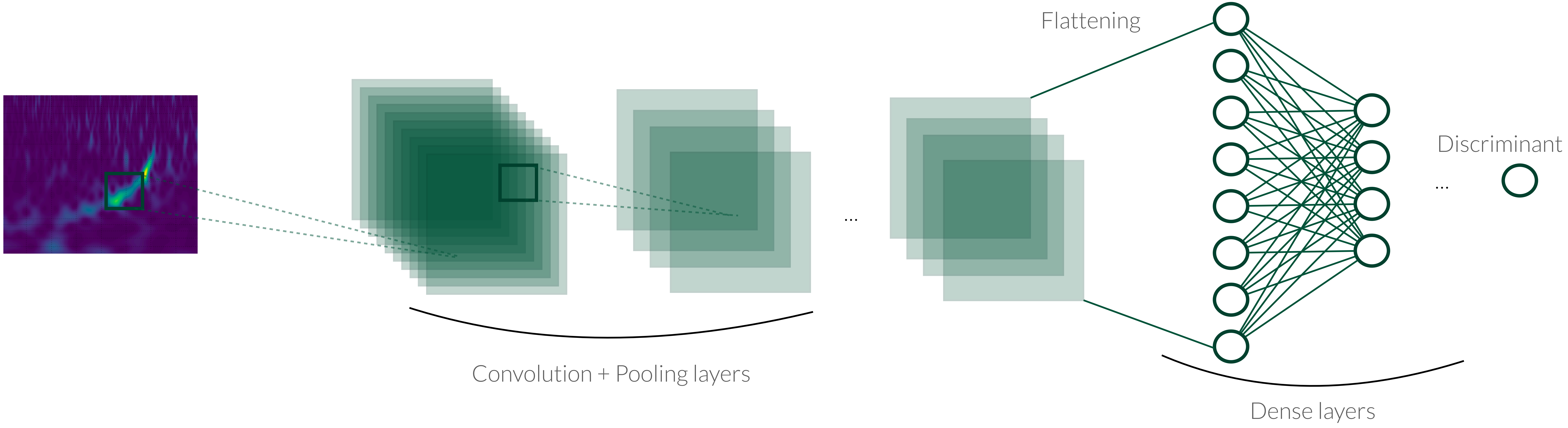
Number of training images:

~ 112.000

Divided equally into signal and real noise (extracted from periods of O3a which do not contain an event)



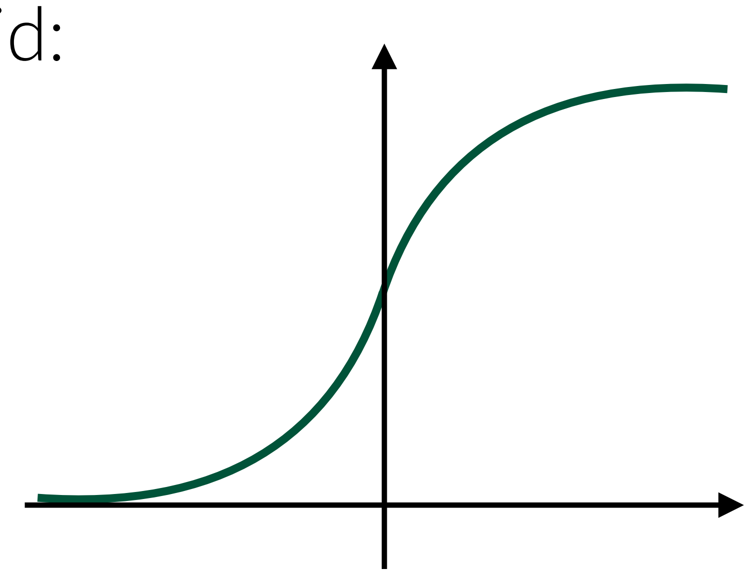
CNN ARCHITECTURE



The actual architecture used is that of a *ResNet50*. For a detailed explanation of the layers see [Phys. Rev. D 103, 062004 (2021)]

The activation function of the last layer is chosen to be a sigmoid:

$$f(x) = \frac{1}{1+e^{-x}}$$

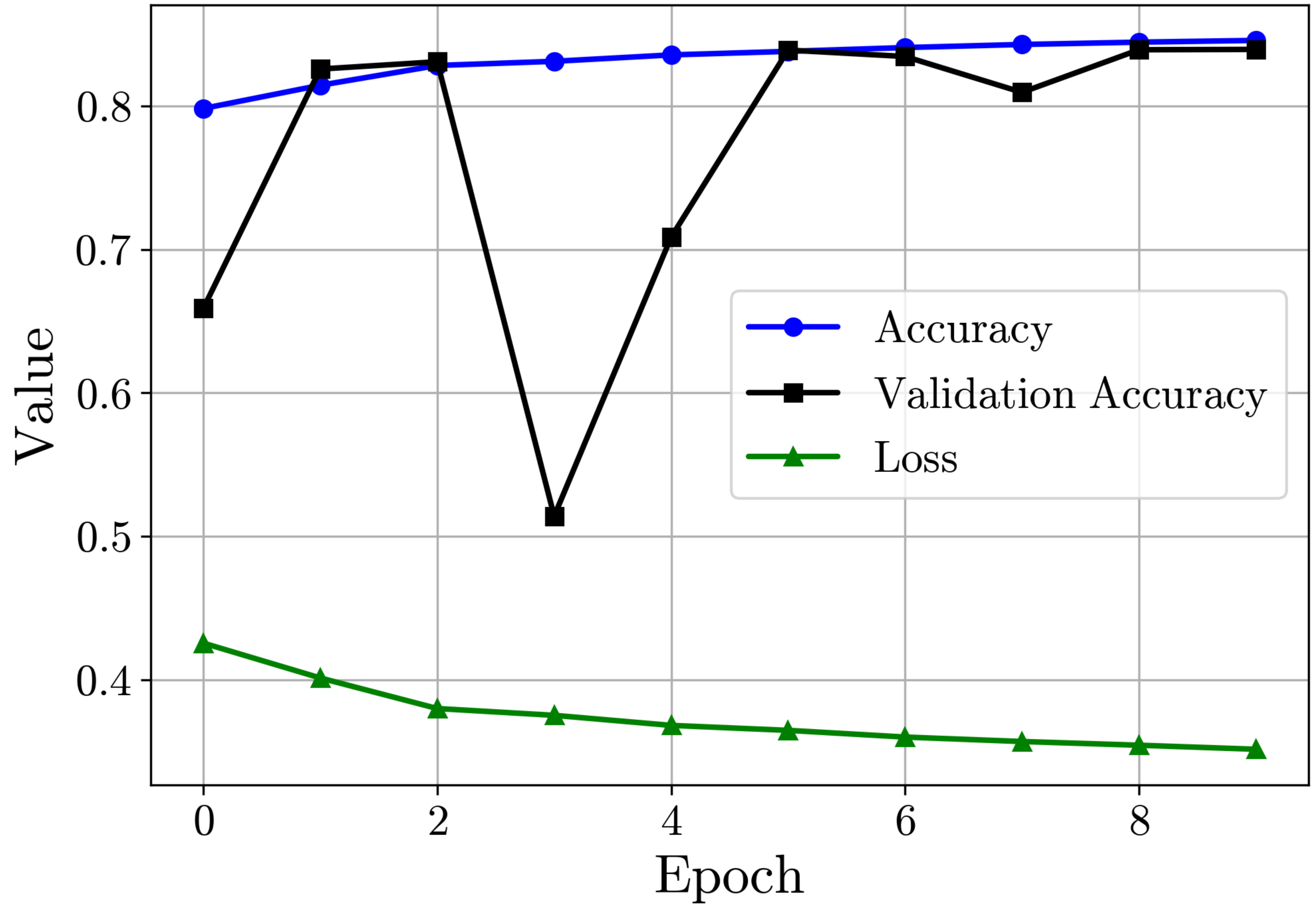


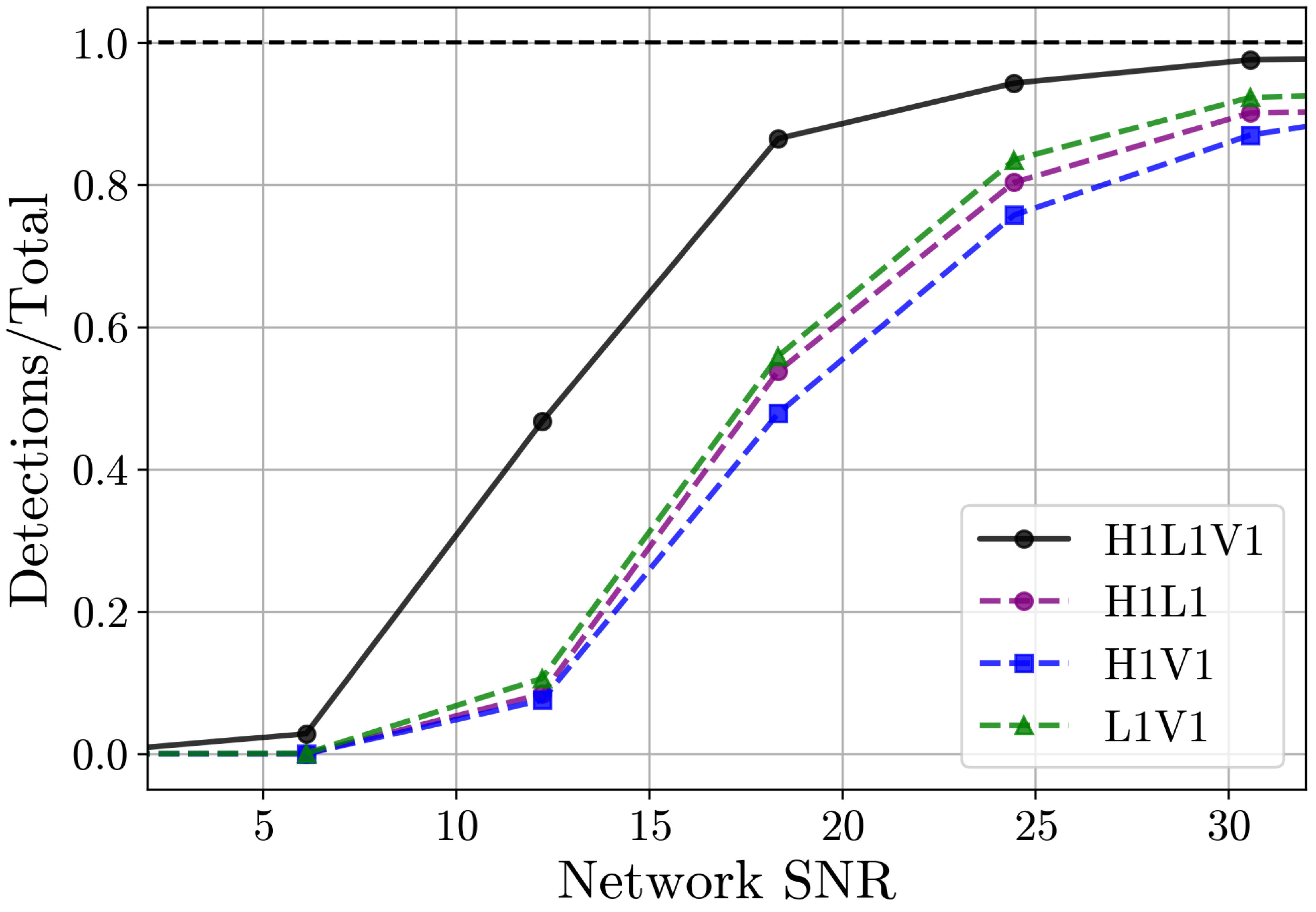
Hyper parameters	
Learning rate	0.001
Mini batch size	32
Maximum number of epochs	12
Optimizer	Adam
Loss function	Binary-cross entropy

Training Results

In total, we have trained 7 CNNs (covering all possible ITF combinations). The metrics that we follow are:

- Loss
- Accuracy
- Validation accuracy (decision metric)





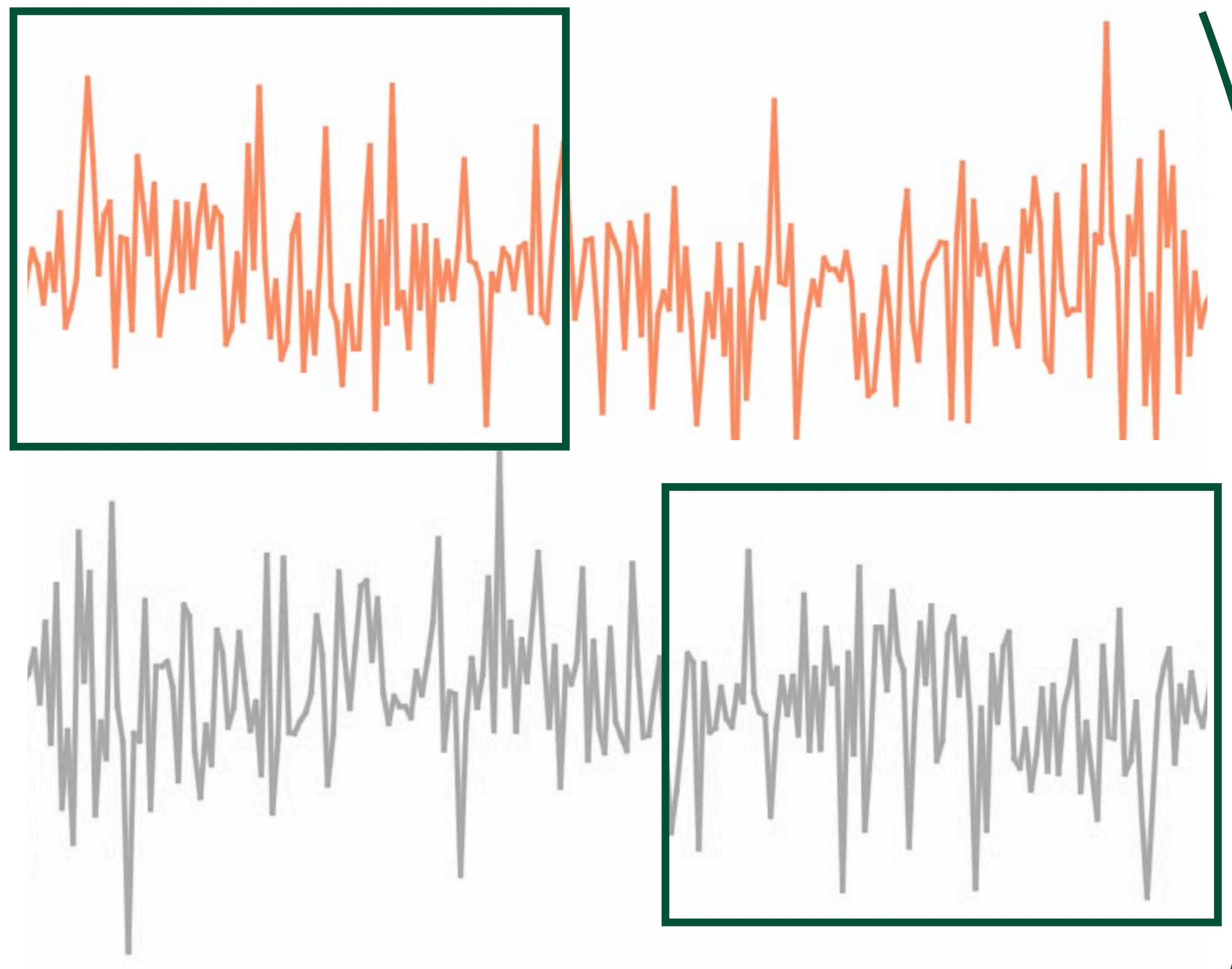
Example for the H1L1V1 case

To establish a threshold to define a trigger we require the discriminant to have associated a False Alarm Rate of $FAR \leq 1\ yrs^{-1}$

To compute the FAR we have performed a time-shifted analysis

ITF 1

ITF 2



Combining these shifted signals we make sure that the final image will be composed of noise and any coincident trigger will be random.

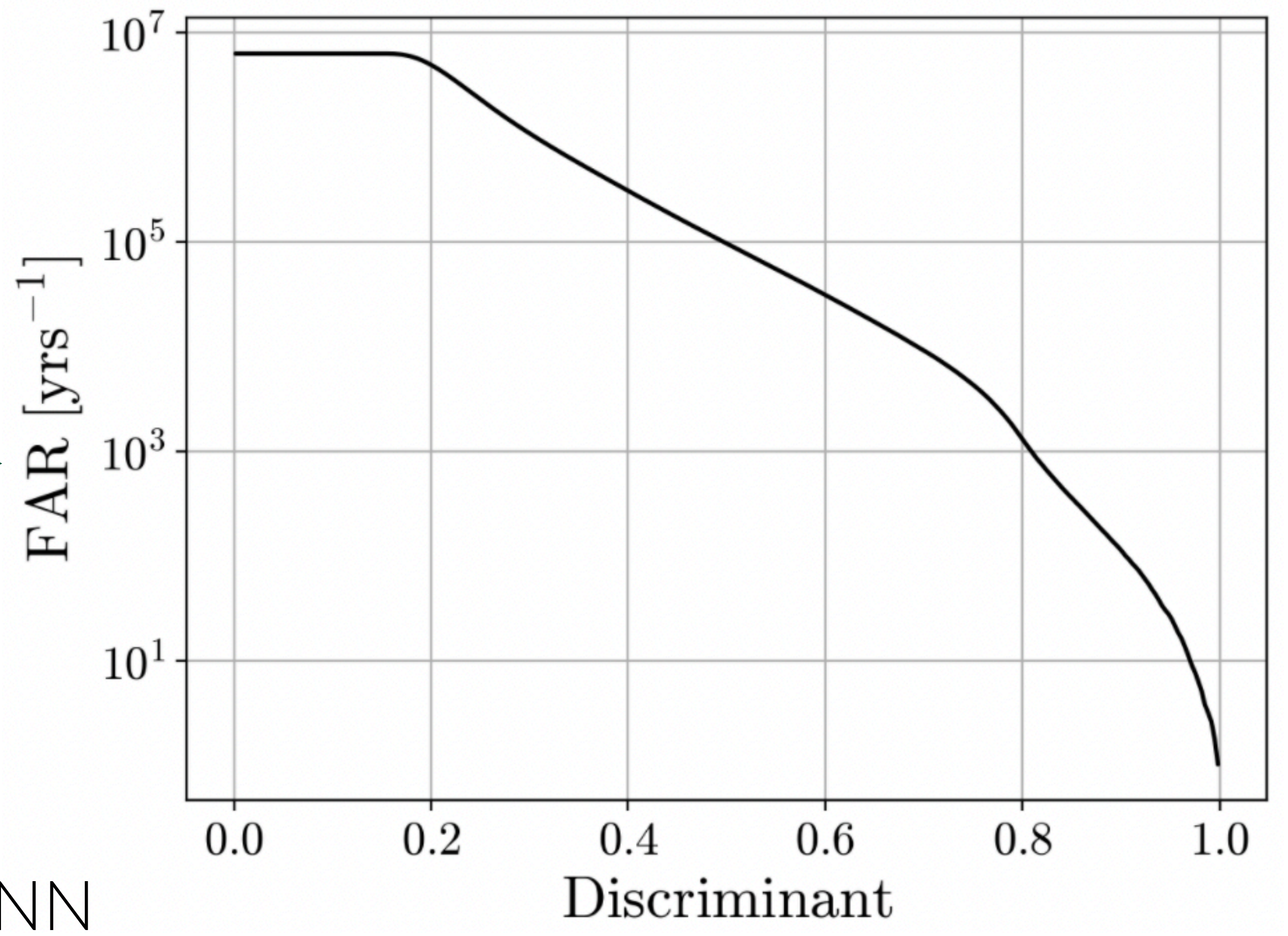
$$FAR(\eta) = \frac{N(\eta)}{T}$$

Number of images with a discriminant equal or higher than η

Discriminant

Total time analyzed

We have analyzed the equivalent of 152 years of data for each period (O3a or O3b) to estimate the FAR. This implies $\sim 10^9$ images processed in total.



O3a CNN

MinoTauro, a GPU cluster held at BSC, was used to analyze the data

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

To enhance the detection of events we have explored the combination of the outputs of the CNNs trained with the information coming from different ITFs.

Method 1 from Physics of the Dark Universe 35 (2022) 100932

$$Loss = \sum_{SNR=5}^{SNR=40} \sigma^2_{SNR} \longrightarrow \sigma^2_{SNR} = \langle D_{SNR} - \langle D_{SNR} \rangle^2 \rangle$$

After running it for our case:

$$D = \beta_1 D_{L1H1V1} + \beta_2 D_{L1H1} + \beta_3 D_{L1V1} + (1 - \beta_1 - \beta_2 - \beta_3) D_{H1V1}$$

$$\beta_1 = 0.03 \quad \beta_2 = 0.30 \quad \beta_3 = 0.33$$

Method 2 from Physics of the Dark Universe 35 (2022) 100932

$$Loss = \sum_{SNR=SNR_0+1}^{SNR=40} \frac{N(D_{SNR} < D_0)}{N_{SNR}} \qquad \frac{N(D_{SNR_0} > D_0)}{N_{SNR_0}} = 0.5 \downarrow D_0$$

SNR_0: SNR at which to accept 50% of the events

After running it for our case: Optimizer didn't converge, the function is not smooth enough

Minimization of the binary cross-entropy

$$Loss = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(D_i) + (1 - y_i) \log(1 - D_i)$$

After running it for our case:

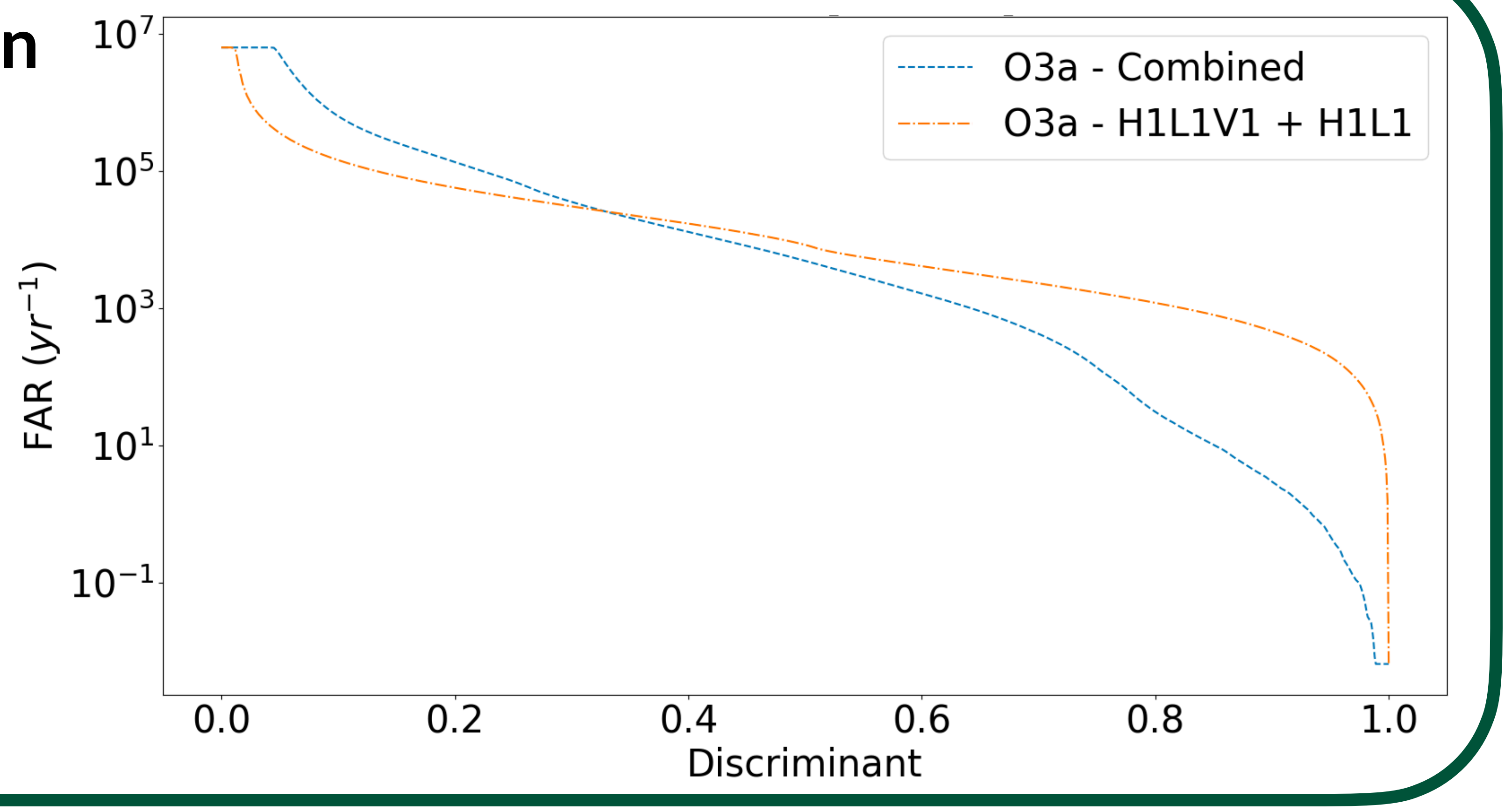
$$D = \beta_1 D_{L1H1V1} + \beta_2 D_{L1H1} + \beta_3 D_{L1V1} + (1 - \beta_1 - \beta_2 - \beta_3) D_{H1V1}$$

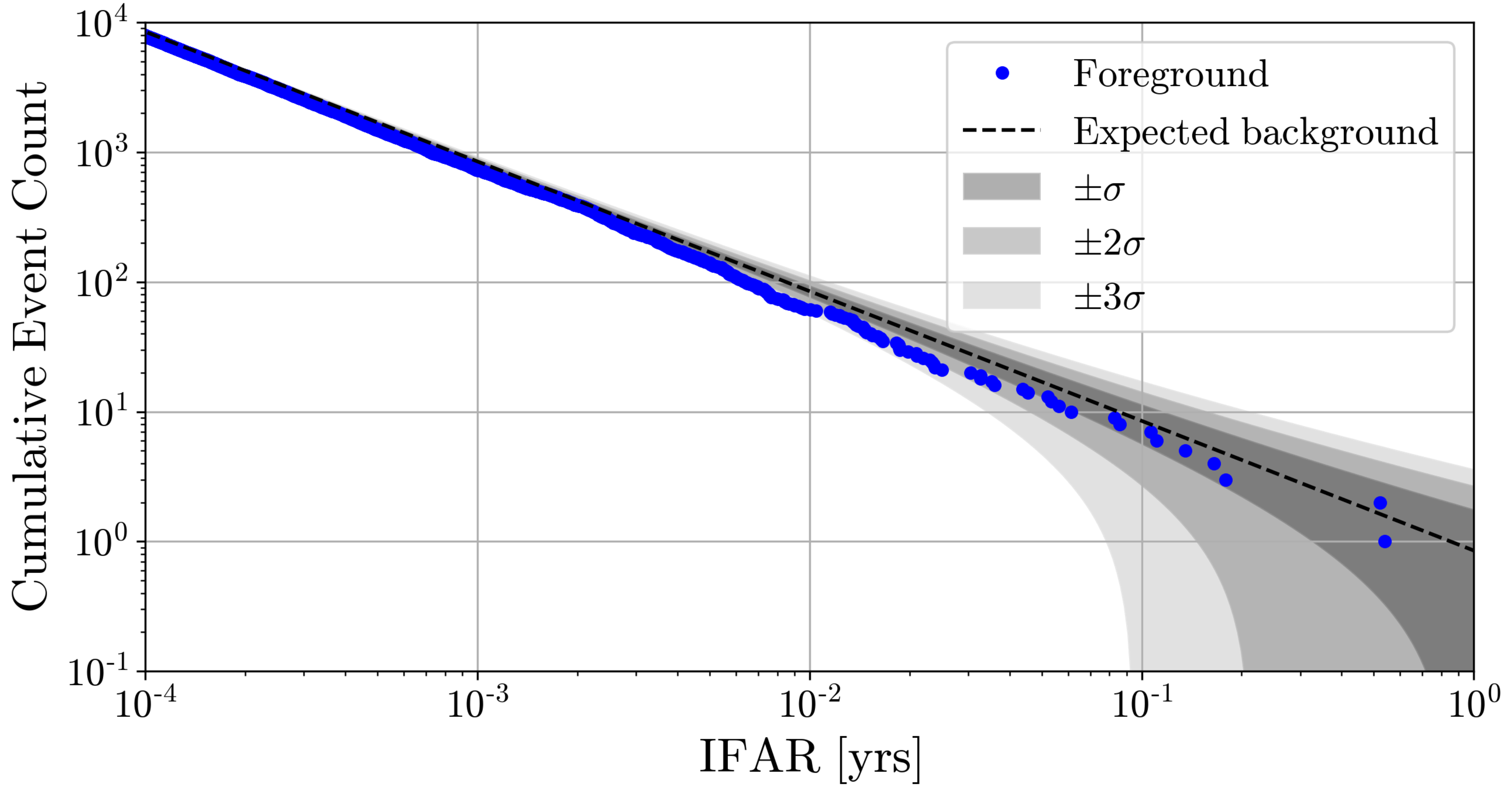
$$\beta_1 = 0.45 \quad \beta_2 = 0.50 \quad \beta_3 = 0.0$$

Arithmetic mean

Gives a good FAR distribution

Example




Cumulative Event Count
IFAR [yrs]
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
10^{-1}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
Foreground
Expected background
$\pm\sigma$
$\pm2\sigma$
$\pm3\sigma$

Implications for the Merger Rate

Effective volume:

$$\langle VT \rangle = \int dz \frac{1}{1+z} \frac{dV_c}{dz} e(z)$$

Comoving volume

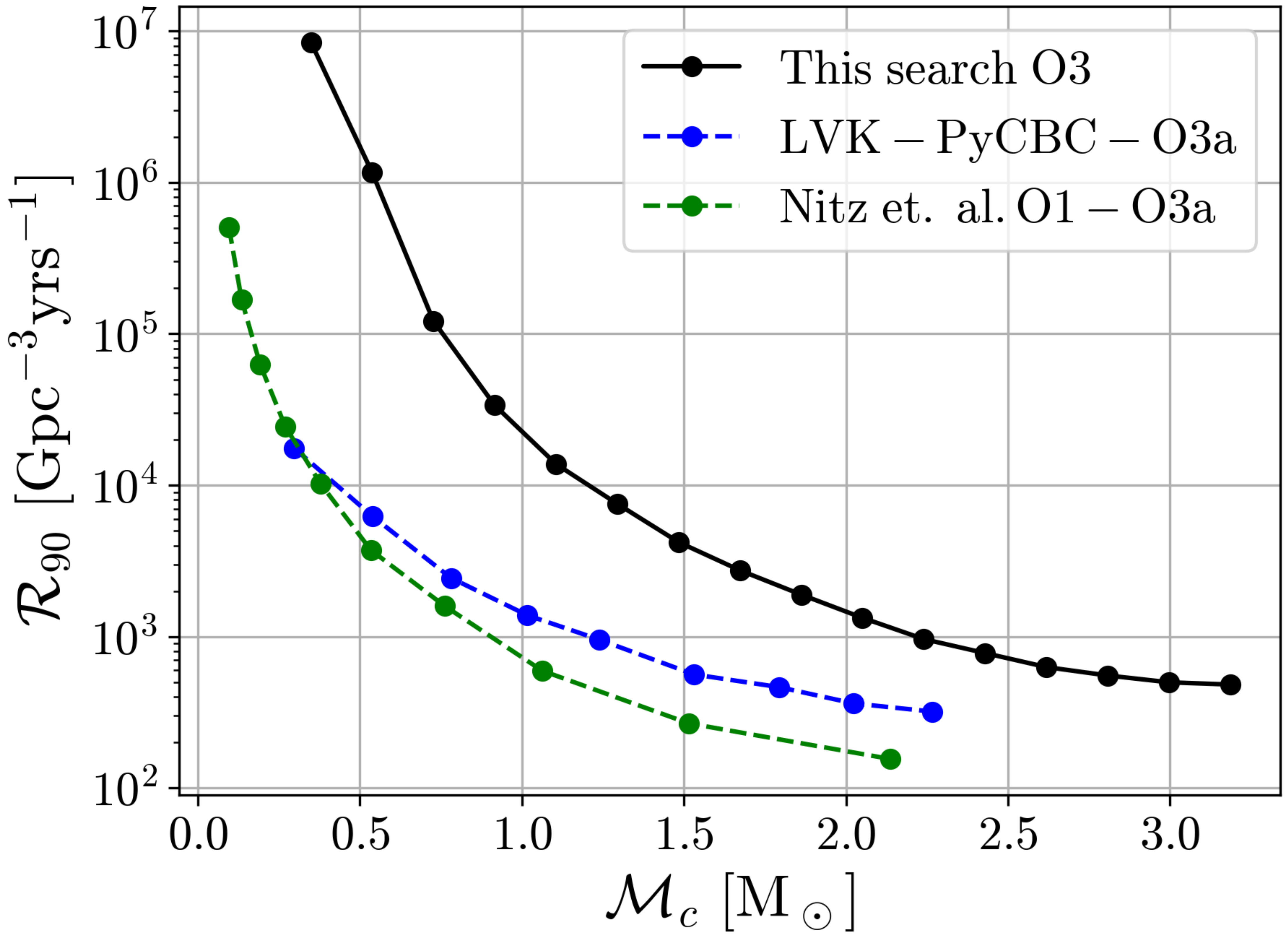
Efficiency (fraction of events that can be recovered)

90% CL upper limit of the merger rate according to the loudest event statistic:

$$\mathcal{R}_{90} = \frac{2.3}{\langle VT \rangle}$$

(See, for example, arXiv:2109.12197)

As expected, we are less sensitive but we cover a broader range.



The next step is to recast this as limits to different theoretical models of Primordial Black Holes

Conclusions

1. We have shown the results using a machine learning approach to detect mass-asymmetric CBC events
2. We have chosen the ResNet50 architecture for excelling in pattern recognition in 2D images
3. This approach is less sensitive than traditional matched filtering but faster (an entire scan takes a day to run instead of months)
4. We have explored a way to combine the outputs of the different CNNs
5. We have been able to establish an upper limit to the merger rate in the mass range considered